BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37206
doi: 10.3897/biss.3.37206

OPEN ACCESS

Conference Abstract

FAIR.ReD: Semantic knowledge graph infrastructure for the life sciences

Lars Vogt[‡], Sören Auer[§], Thomas Bartolomaeus[‡], Pier Luigi Buttigieg[|], Peter Grobe[¶], Peter Michalik [#], Markus Stocker[§], Ricardo Usbeck[¤]

‡ Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany
§ Technische Informationsbibliothek (TIB), Hannover, Germany
| Alfred-Wegener-Institut für Polar und Meeresforschung, Bremen, Germany
¶ Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany
# Zoologisches Institut & Museum, Greifswald, Germany
¤ Fraunhofer Institut für Intelligente Analyse- und Informationssysteme (IAIS), München, Germany

Corresponding author: Lars Vogt (lars.m.vogt@googlemail.com)

Received: 12 Jun 2019 | Published: 19 Jun 2019

Citation: Vogt L, Auer S, Bartolomaeus T, Buttigieg P, Grobe P, Michalik P, Stocker M, Usbeck R (2019) FAIR.ReD: Semantic knowledge graph infrastructure for the life sciences. Biodiversity Information Science and Standards 3: e37206. https://doi.org/10.3897/biss.3.37206

Abstract

We would like to present *FAIR Research Data: Semantic Knowledge Graph Infrastructure for the Life Sciences* (in short, FAIR.ReD), a project initiative that is currently being evaluated for funding. FAIR.ReD is a software environment for developing data management solutions according to the FAIR (Findable, Accessible, Interoperable, Reusable; Wilkinson et al. 2016) data principles. It utilizes what we call a Data Sea Storage, which employs the idea of Data Lakes to decouple data storage from data access but modifies it by storing data in a semantically structured format as either semantic graphs or semantic tables, instead of storing them in their native form. Storage follows a top-down approach, resulting in a standardized storage model, which allows sharing data across all FAIR.ReD Knowledge Graph Applications (KGAs) connected to the same Sea, with newly developed KGAs having automatically access to all contents in the Sea. In contrast access and export of data follows a bottom-up approach that allows the specification of additional data models to meet the varying domain-specific and programmatic needs for accessing structured data. The FAIR.ReD engine enables bidirectional data conversion between the two storage models and any additional data model, which will substantially reduce

© Vogt L et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

conversion workload for data-rich institutes (Fig. 1). Moreover, with the possibility to store data in semantic tables, FAIR.ReD provides high performance storage for incoming data streams such as sensory data.

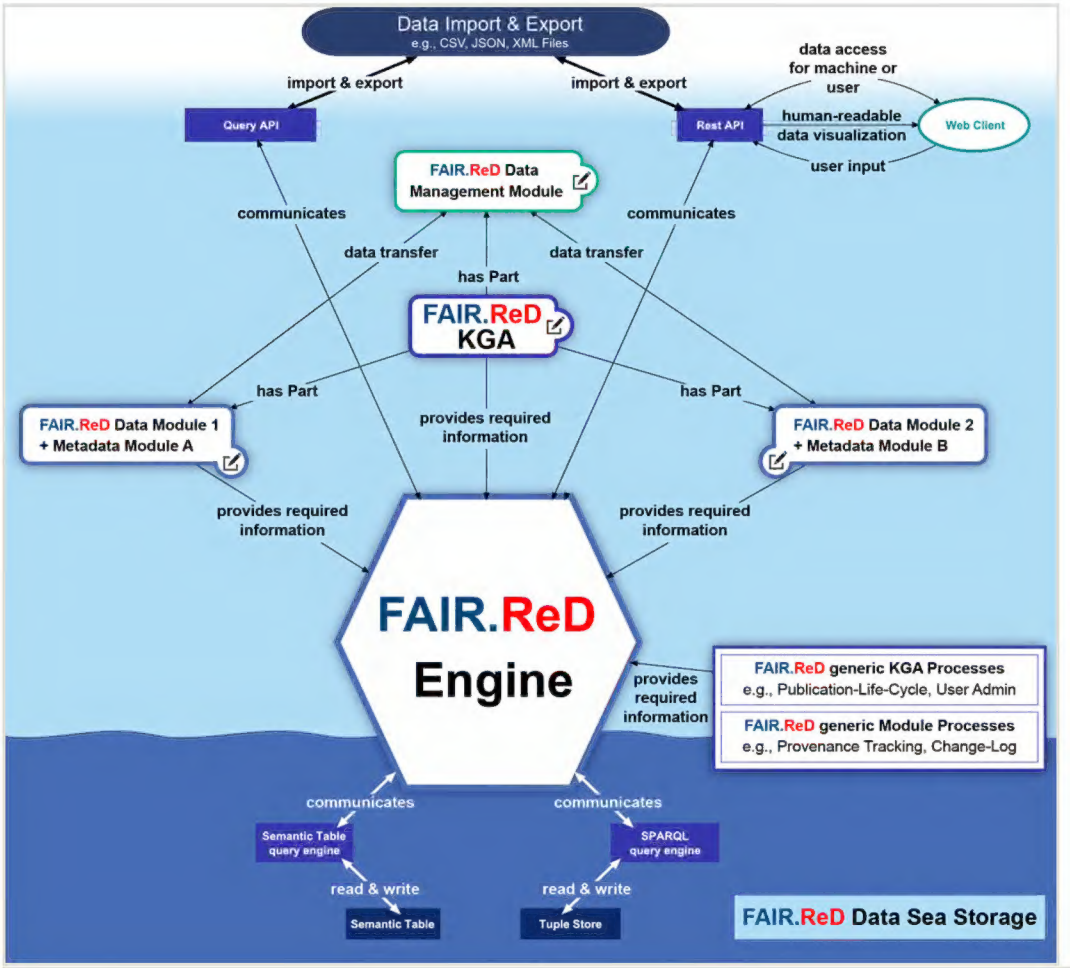

Figure 1.

Example of a FAIR.ReD KGA that uses two FAIR.ReD modules. Schema of the general architecture of a FAIR.ReD KGA, with two FAIR.ReD data modules and their accompanying FAIR.ReD metadata modules and the FAIR.ReD Data Sea Storage. FAIR.ReD modules and KGAs are editable using the FAIR.ReD editor. All unpublished contents (i.e., contents that are still in draft status) in the FAIR.ReD Data Sea Storage are not openly accessible and require access rights.

FAIR.ReD KGAs are modularly organized. Modules can be edited using the FAIR.ReD editor and combined to form coherent KGAs. The editor allows domain experts to develop their own modules and KGAs without any programming experience required, thus also allowing smaller projects and individual researchers to build their own FAIR data management solution.

Contents from FAIR.ReD KGAs can be published under a Creative Commons license as documents, micropublications, or nanopublications, each receiving their own DOI. A publication-life-cycle is implemented in FAIR.ReD and allows updating published contents for corrections or additions without overwriting the originally published version. Together with the fact that data and metadata are semantically structured and machine-readable, all contents from FAIR.ReD KGAs will comply with the FAIR Guiding Principles. Due to all FAIR.Red KGAs providing access to semantic knowledge graphs in both a human-readable and a machine-readable version, FAIR.ReD seamlessly integrates the complex

RDF (Resource Description Framework) world with a more intuitively comprehensible presentation of data in form of data entry forms, charts, and tables.

Guided by use cases, the FAIR.ReD environment will be developed using semantic programming where the source code of an application is stored in its own ontology. The set of source code ontologies of a KGA and its modules provides the steering logic for running the KGA. With this clear separation of steering logic from interpretation logic, semantic programming follows the idea of separating main layers of an application, analog to the separation of interpretation logic and presentation logic. Each KGA and module is specified exactly in this way and their source code ontologies stored in the Data Sea. Thus, all data and metadata are semantically transparent and so is the data management application itself, which substantially improves their sustainability on all levels of data processing and storing.

Keywords

semantic knowledge graph; Data Sea Storage; FAIR data principle; semantic programming; knowledge graph application

Presenting author

Lars Vogt

Presented at

Biodiversity_Next 2019

References

- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 https://doi.org/10.1038/sdata.2016.18